A conditional extreme value volatility estimator based on high-frequency returns

Turan G. Bali\textsuperscript{a,}\textsuperscript{*}, David Weinbaum\textsuperscript{b}

\textsuperscript{a}Department of Economics & Finance, Baruch College, Zicklin School of Business, City University of New York, 17 Lexington Avenue, Box 10-225, NY 10010, USA
\textsuperscript{b}Department of Finance, Johnson Graduate School of Management, Cornell University, 375 Sage Hall, Ithaca, NY 14853, USA

Received 7 July 2004; accepted 2 October 2005
Available online 10 March 2006

Abstract

This paper introduces a conditional extreme value volatility estimator (EVT) based on high-frequency returns. The relative performance of the EVT is compared with the discrete-time GARCH and implied volatility models for 1-day and 20-day-ahead forecasts of realized volatility. This is also a first attempt towards detecting any time-series variation in extreme value distributions using high-frequency intraday data. The information content of EVT is examined in the context of forecasting S&P 100 index volatility. Adjusted-$R^2$ values imply superior performance of the implied volatility index (VIX) and EVT in capturing time-series variation in realized volatility. The forecasting ability of various discrete-time GARCH models turns out to be inferior to VIX and EVT. According to the Theil inequality coefficient and the heteroscedasticity-adjusted root mean squared and mean absolute errors, (1) EVT provides more accurate forecasts than the VIX and GARCH models; (2) VIX generally yields a less accurate characterization of realized volatility than EVT and GARCH models. These results have implications for financial risk management, and are thus relevant to both regulators and practitioners.

© 2006 Elsevier B.V. All rights reserved.

\textit{JEL classification:} G12; C13; C22

\textit{Keywords:} Extreme value; Realized volatility; High-frequency returns; GARCH; Implied volatility

\*Corresponding author. Tel.: +646 312 3506; fax: +646 312 3451.
E-mail addresses: turan_bali@baruch.cuny.edu (T.G. Bali), dw85@cornell.edu (D. Weinbaum).
1. Introduction

Modeling and estimating the volatility of economic time series has been high on the agenda of financial economists over the last two decades. Engle (1982) put forward the Autoregressive Conditional Heteroskedastic (ARCH) class of models for conditional variances, which proved to be extremely useful for analyzing financial return series. Since then an extensive literature has been developed for modeling the conditional distribution of stock prices, interest rates, and exchange rates.

Although many studies show that the parameters of different ARCH models are highly significant in-sample, there is mixed evidence that they provide good out-of-sample forecasts of equity return volatility. Some studies [e.g., Akgiray (1989), Frances and Van Dijk (1995), Brailsford and Faff (1996), and Figlewski (1997)] examine the out-of-sample predictive ability of ARCH models, and find that a regression of realized volatility on estimated volatility yields a low $R^2$ statistic. Andersen and Bollerslev (1998a, b) prove that regression methods will produce low $R^2$ values when realized volatility is measured by daily squared returns, even for optimal GARCH forecasts, because squared daily returns are noisy estimates of volatility. They show that intraday returns can be used to construct a realized volatility series that reduces the noise in measurements of integrated volatility. They find remarkable improvements in the forecasting performance of ARCH models when they are used to forecast the new realized series, compatible with theoretical analysis.

An alternative approach to GARCH volatility forecasts is to use implied volatilities from options. Day and Lewis (1992) and Lamoureux and Lastrapes (1993) examine implied volatility (IV) as a source of information. Both studies find that IV contributes a statistically significant amount of information about volatility over the (short-term) forecasting horizon covered by the models, but they also find that IV does not fully impound the information that the model is able to extract from historical prices. Day and Lewis conclude that IV performs as well but no better than forecasts from ARCH models, and mixtures of two forecasts outperform both univariate forecasts. Lamoureux and Lastrapes examine forecasting volatility through option expiration and find that IV alone is less accurate than the models that incorporate historical prices in eight out of 10 cases. Canina and Figlewski (1993) provide contrary evidence, and indicate that implied volatilities are poor forecasts of volatility, and simple historical volatilities outperform implied volatilities. Christensen and Prabhala (1998) show for a much longer period that while implied volatilities are biased forecasts of volatility they perform better than historical volatility models. Fleming (1998) also provides evidence that implied volatilities are more informative than daily returns when forecasting equity volatility.

The importance of intraday returns for measuring realized volatility is demonstrated by Andersen and Bollerslev (1998a, b), Andersen et al. (2001a, b, 2003), Barndorff-Nielsen and Shephard (2001, 2002a, b), and Andreou and Ghysels (2002). They suggest the sum of squared high-frequency intraday returns as an
approximation to the daily volatility. This quadratic variation is referred to as an estimator of integrated volatility. Blair et al. (2001) explore the incremental volatility information of high-frequency (5-min) stock index returns. They answer some important empirical questions for the S&P 100 index. Based on their in-sample analysis of low-frequency (daily or weekly) data using GARCH models, Blair et al. find no evidence for incremental information in daily index returns beyond that provided by the implied volatility index (VIX). Their extension of the historical information set to include high-frequency returns suggests that there is only minor incremental information in high-frequency returns, and this information is almost subsumed by implied volatilities. Their out-of-sample comparisons of volatility forecasts show that VIX provides more accurate forecasts than either low- or high-frequency index returns, regardless of the definition of realized volatility and the horizon of the forecasts.

Volatility is a central concept in finance, whether in derivative pricing, asset allocation, or risk management. Despite the importance of conditional volatility, the existing literature has not yet reached an agreement on whether implied, GARCH or stochastic volatility estimators provide more accurate forecasts of realized volatility.

This paper introduces a conditional extreme value volatility estimator (EVT) based on high-frequency returns, and compares its performance with implied and GARCH volatility models for 1-day and 20-day-ahead forecasts of realized volatility. The information content of the EVT is examined, and several important questions for forecasting stock market volatility are answered. The relative performance of discrete-time GARCH, VIX, and EVT estimators for 1-day and 20-day-ahead forecasts of realized volatility is found to be sensitive to the choice of performance measures. The adjusted-\(R^2\) values imply superior performance of the VIX and EVT in capturing time-series variation in realized volatility. The forecasting ability of discrete-time GARCH models with alternative distribution functions turns out to be inferior to VIX and EVT. In fact, there appears to be no incremental information in GARCH estimates that adds to the explanation of realized volatility. Although the adjusted-\(R^2\) provides the direction and magnitude of the relationship between the realized and estimated volatilities, it does not measure how far the volatility forecasts are away from the realized standard deviation. When the summary statistics are calculated based on the deviation between forecasts and realizations, the relative performance of alternative volatility models changes. According to the Theil inequality coefficient, the heteroscedasticity-adjusted root mean squared and mean absolute errors, EVT provides more accurate forecasts than the VIX and GARCH volatility models, and VIX generally yields less accurate characterization of realized volatility than EVT and GARCH models.

Several important results emerge from our analysis. First, we show that it is possible to construct good volatility forecasts based solely on intraday extreme

---

1We apply the extreme value theory (hence the acronym EVT) based approach to S&P 100 index returns to compare the empirical performance of EVT with an implied volatility index (VIX). The same approach can easily be applied to any financial asset returns including exchange rate, interest rate, equity, commodity, and futures returns.
returns. In fact, the information content of intraday extreme returns is such that the resulting volatility model fares well when compared to the traditional stalwarts (e.g., IV and GARCH). Perhaps even more surprising is the fact that EVT compares favorably to forecasts derived directly from realized volatility (as in ARMA-type models), which aggregate squared intraday returns and therefore, by construction, exploit the information in the entire series of intraday returns. Of particular practical interest, EVT is easily implemented using standard maximum likelihood methods, and is thus a useful tool for measuring and forecasting volatility.

Second, we show that standard extreme value theory, which routinely assumes iid data, can be useful in analyzing asset returns, which are, of course, conditionally heteroskedastic and thus not i.i.d. We accomplish this (i) by focusing on per-period (daily) extremes, as opposed to exceedances over high thresholds, and (ii) by modeling explicitly the dependence structure in the per-period conditional extreme value distributions. As a byproduct of our analysis, we find that there is in fact substantial time series variation in daily extreme value distributions using intraday data.

Third, our analysis has implications for financial risk management, and is thus relevant to both regulators and practitioners. In recent years, a central issue in risk management has been to determine capital requirements for financial institutions to meet catastrophic market risk. The standard volatility models can be successful in estimating the maximum likely loss of an institution under normal market conditions. However, as pointed out by Longin (2000) and Bali (2003), the traditional volatility measures based on the distribution of all returns cannot produce accurate estimates of market risks during highly volatile periods. Longin and Bali introduce an unconditional extreme value approach to calculating value at risk (VaR). Although their risk measurement technique is based on a sound statistical theory, it does not yield VaR measures which reflect the current volatility background. Given the serial correlation and conditional heteroscedasticity of most financial data, the use of unconditional volatility is a major drawback of any kind of VaR estimator. The conditional methods developed in this paper can be readily applied to extend and improve the unconditional extreme value risk management approach of Longin and Bali. This is important because financial institutions calculate and monitor their maximum likely loss on a daily basis, and a daily VaR can be estimated easily using the conditional extreme value approach proposed in the paper.2

The paper is organized as follows. Section 2 provides alternative volatility models. Section 3 describes the data on intraday and daily index returns, and daily realized, implied, and GARCH volatilities. Section 4 presents the estimation results. Section 5 discusses the in-sample and out-of-sample performance of implied, GARCH, and

---

2Funds managers report two types of realized returns. One is average quarterly realized returns, and the other reported by hedge funds often on demand is the maximum drawdown, which is in fact the extreme loss for the life of the hedge fund. Thus, the conditional distribution of extreme returns will carry additional information about the risk taken by hedge fund managers.
EVTs. Section 6 provides extensions and robustness of our findings. Section 7 concludes the paper.

2. Alternative volatility models

The empirical performance of alternative volatility forecasts requires a measure of volatility realizations to assess fit. While the traditional approach of using squared daily returns for this purpose has been criticized in the literature [see, e.g., Andersen and Bollerslev (1998a)], one could imagine applications in which this would indeed be the quantity of interest. In this paper, we assess fit with respect to both measures of volatility: (i) realized or integrated volatility (sum of squared high frequency returns), and (ii) the traditional measure based upon squared daily returns.

2.1. Implied volatility

To calculate an IV, an option valuation model (such as the Black-Scholes (1973) model) is needed as well as inputs for that model (price of the underlying asset, exercise price, risk-free rate, time-to-expiration, dividends) and an observed option price. Many of these variables are subject to measurement errors that may induce biases in a series of implied volatilities.

We use the CBOE VIX to mitigate the problems caused by the use of an inappropriate option valuation model [see Harvey and Whaley (1992)] and infrequent trading of the stocks in the index [see Jorion (1995)]. VIX is constructed so that it represents a hypothetical option that is at-the-money and has a constant 22 trading days (30 calendar days) to expiration. VIX is a weighted index of American implied volatilities calculated from eight near-the-money, near-to-expiry, S&P 100 call and put options and it is constructed so as to reduce mismeasurement and smile effects. This makes it a more accurate measurement of implied market volatility than IV from any single option.\(^3\),\(^4\)

2.2. Discrete-time GARCH models

Following the introduction of ARCH processes by Engle (1982) and their generalization by Bollerslev (1986), there have been numerous refinements of this approach to modeling conditional volatility.\(^5\) This paper uses the linear symmetric GARCH model of Bollerslev (1986) and the absolute value GARCH process of

---

3For a detailed explanation of the construction of the VIX index, see Whaley (1993) and Fleming et al. (1995).

4Our empirical analysis uses the original version of VIX. The new VIX, introduced by the CBOE on September 22, 2003, is obtained from S&P 500 index option prices and incorporates information from the volatility skew by using a wider range of strike prices rather than just at-the-money series.

Taylor (1986) and Schwert (1989) to estimate time-varying conditional volatility. Most asset pricing models postulate a positive relationship between stock market expected return and risk. The following GARCH-in-mean process is used with alternative density functions to model the relation between mean returns on a stock portfolio and its conditional volatility:

\[ R_t = x_0 + x_1 \sigma_{t|t-1} + z_t \sigma_{t|t-1} = \mu_{t|t-1} + z_t \sigma_{t|t-1}, \]  

\[ GARCH : \quad \sigma_{t|t-1}^2 = \beta_0 + \beta_1 z_{t-1}^2 \sigma_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \]  

\[ TS-GARCH : \quad \sigma_{t|t-1} = \beta_0 + \beta_1 |z_{t-1}| \sigma_{t-1} + \beta_2 \sigma_{t-1}, \]

where \( R_t \) is daily return on S&P 100 index for period \( t \) defined in the standard way by the natural logarithm of the ratio of consecutive daily closing index levels. \( \mu_{t|t-1} = x_0 + x_1 \sigma_{t|t-1} \) is the conditional mean for period \( t \) based on the information set up to time \( t-1 \), \( \Omega_{t-1} \). \( \sigma_{t|t-1}^2 \) and \( \sigma_{t|t-1} \) are the conditional variance and standard deviation of daily returns based on \( \Omega_{t-1} \). In Eqs. (2) and (3), the current conditional volatility is defined as a function of the last period’s unexpected news (or information shocks), \( z_{t-1} \), and the last period’s volatility, \( \sigma_{t-1} \).

Most empirical work shows that the excess kurtosis values for financial returns are extremely high and statistically significant, implying that the tails of the empirical distribution are much thicker than the tails of the normal distribution. In light of the empirical evidence of fat-tailed errors, Bollerslev (1987) and Nelson (1991) use leptokurtic distributions such as the standardized \( t \) and the generalized error distribution (GED). In this paper we use the fat-tailed conditional GED density given by:

\[ GED : \quad f(R_t; \mu_{t|t-1}, \sigma_{t|t-1}, v) = \frac{v}{\lambda} \exp\left(\frac{-(1/2)|z_t/\lambda|^v}{2\Gamma(1/v)}\right), \]

where

\[ z_t = \frac{R_t - \mu_{t|t-1}}{\sigma_{t|t-1}} , \quad \lambda = \left[ \frac{2(-2/v) \Gamma(1/v)}{\Gamma(3/v)} \right]^{1/2} , \quad \Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx \]

is the gamma function, and \( v \) is a positive parameter, or degrees of freedom governing the thickness of the tails.\(^8\)

\(^6\)At an earlier stage of the study, the asymmetric GARCH models of Engle and Ng (1993), Glosten et al. (1993), and Zakoian (1994) were used also. To save space, we do not present the empirical results from the NGARCH, GJR-GARCH, and TGARCH models. They are very similar to those reported in our tables, and available upon request.

\(^7\)At an earlier stage of the study, we used the fat-tailed GED and Student-\( t \) as well as the thin-tailed normal distribution. Since the results turn out to be similar, and the GED TS-GARCH model performs slightly better than alternative GARCH specifications, we report results from the GED TS-GARCH model. The full set of details is available upon request.

\(^8\)For \( v = 2 \), the GED yields the normal distribution, while for \( v = 1 \) it yields the Laplace or the double exponential distribution. If \( v < 2 \), the density has thicker tails than the normal, whereas for \( v > 2 \) it has thinner tails.
2.3. A conditional extreme value volatility estimator

This section introduces a new approach that uses only the extremes in the stock price process to analyze the dynamic behavior of financial return volatility. The key insight is that volatility dynamics can be directly related to the changing parameters of extremal distributions. That is to say, volatility dynamics can be captured to a large extent by a parsimonious extremal distribution theory, albeit with time-varying parameters.9

As pointed out by Diebold et al. (1998) and Diebold and Schuermann (1999), a pitfall of using the standard extreme value theory in financial applications is that it is developed almost exclusively for iid series, whereas financial return series are serially correlated and heteroskedastic, and thus not iid. Following Diebold et al. (1998), we focus on per-period extremes, and not on exceedances over thresholds. Given that our objective is to construct conditional daily volatility forecasts, we take the basic period to be one day, and model the extreme intraday returns within the day. Of importance, we do not assume that the daily extremes (e.g., the largest 5-min returns each day) are iid; rather, we model the dependence in daily extremes explicitly.

To set forth notation, let \( \ln P_t \) denote the time \( t \) log level of the S&P 100 index, with the unit interval corresponding to one day. The discretely observed time series process of index returns with \( q \) observations per day, or a return horizon of \( 1/q \), is then defined by

\[
R_q(t) = \ln P_t - \ln P_{t-1/q},
\]

where \( t = \{1/q, 2/q, \ldots\} \). In our empirical analysis, we use 5-min returns (\( q = 79 \)), 15-min returns (\( q = 26 \)), and 30-min returns (\( q = 13 \)) to check the robustness of our results.10

Let \( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \) be a sequence of 5-min returns on intraday \( q \) = 1/79, 2/79, 3/79, …, 79/79 on day \( t \). Extremes are defined as the maxima and minima of the \( q \) random variables \( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \).

Let \( M_{q,t} \) represent the highest (maximum) and \( m_{q,t} \) denote the lowest (minimum) 5-min returns over trading day \( t \):

\[
M_{q,t} = \max\left( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \right),
\]

\[
m_{q,t} = \min\left( R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t} \right).
\]

---

9The extreme value distributions have recently become popular in the finance literature because of their superior performance in estimating VaR, see e.g. Longin (2000) and Bali (2003) for the risk management performance of unconditional extreme value distributions. See e.g. McNeil and Frey (2000) for a conditional approach to VaR.

10We present our results from 5-min returns because the daily maxima and minima are obtained from the intraday data, and we prefer to generate extremes from a larger intraday sample. Using the highest frequency provides larger intraday sample, and yields more reliable extreme observations [Andersen et al. (2000) use the volatility signature plots to determine the optimal sampling frequency. Fleming et al. (2003) provide a simple bias correction]. Note that the empirical findings from 15- and 30-min returns on S&P 100 index turn out to be very similar to those reported in our tables. The parameter estimates and performance measures of extreme value volatility estimators based on 15- and 30-min returns are available upon request.
\[ m_{q,t} = \min(R_{(1/79),t}, R_{(2/79),t}, R_{(3/79),t}, \ldots, R_{(79/79),t}). \]  

To find a limit distribution for maxima, the maximum variable, \( M_{q,t} \), is transformed such that the limit distribution of the new variable is a non-degenerate one. Following Fisher-Tippett (1928) theorem, the variate, \( M_{q,t} \), is reduced with a location parameter, \( \mu_{q,t} \), and a scale parameter, \( \sigma_{q,t} \), in such a way that

\[
x_t = (M_{q,t} - \mu_{q,t})/\sigma_{q,t} \xrightarrow{d} H_{\text{max}(x)}.
\]

Assuming the existence of a sequence of such coefficients \{\( \mu_{q,t}, \sigma_{q,t} > 0 \}\}, three types of non-degenerate distributions are obtained for the standardized maximum:

- **Frechet**: \[ H_{\text{max},\xi}(x) = \exp(-x^{-1/\xi}) \],

- **Weibull**: \[ H_{\text{max},\xi}(x) = \exp(-(-x)^{-1/\xi}) \],

- **Gumbel**: \[ H_{\text{max},0}(x) = \exp[- \exp(-x)] \].

Jenkinson (1955) proposes a generalized extreme value (GEV) distribution, which includes the three limit distributions in (8)–(10), distinguished by Gnedenko (1943):

\[
H_{\text{max}}(M; \mu, \sigma, \xi) = \exp\left\{ -\left[ 1 + \xi \left( \frac{M - \mu}{\sigma} \right) \right]^{-1/\xi} \right\},
\]

\[
1 + \xi \left( \frac{M - \mu}{\sigma} \right) \geq 0,
\]

where \( \xi \) is a shape parameter. For \( \xi > 0 \), \( \xi < 0 \), and \( \xi = 0 \) we obtain the Frechet, Weibull and Gumbel families, respectively. The Frechet distribution is fat tailed as its tail is slowly decreasing; the Weibull distribution has no tail—after a certain point there are no extremes; the Gumbel distribution is thin-tailed as its tail is rapidly decreasing. The shape parameter \( \xi \), called the tail index, reflects the fatness of the distribution (i.e., the weight of the tails), whereas the parameters of scale, \( \sigma \), and of location, \( \mu \), determine the mean and standard deviation of extremes along with \( \xi \).

One potential problem with Eqs. (8)–(11) is that the 79 intraday returns are not i.i.d. As explained by Leadbetter et al. (1983), Resnick (1987) and Castillo (1988), when the data are dependent, it is possible that extreme value distributions cannot be described as in (8)–(11). Not surprisingly, no general statement can be made without further assumptions on the exact nature of the dependence structure. However, the theory of extremes for the case of dependence, while not completely developed, has identified a number of empirically relevant cases for which inferences based on standard extreme value theory remain valid. For example, this is true if the intraday returns are stationary, and follow an \( MA(q) \), \( AR(p) \), or \( ARMA(p,q) \) model (see Leadbetter et al. (1983), Resnick (1987) and Castillo (1988) for detailed treatment). Because these conditions describe our data reasonably well, we prefer to stick to the extreme value distributions described in (8)–(11) as opposed to making assumptions on the dependence structure that would lead to other forms of extreme value distributions.
To make sure that our results are not driven by a misspecified conditional extreme value distribution, we follow Diebold et al. (1998) and conduct the following robustness check: we first fit a conditional mean-volatility model to the raw intraday returns, standardize the data by the estimated conditional mean and volatility, and then repeat our analysis based on these standardized residuals. The results (presented in Section 6) remain qualitatively the same and highlight the superior performance of EVT.

Our main objective is to estimate the conditional volatility of extremes based on high-frequency intraday data. Appendix A presents the first and second moments of the maxima and minima for the GEV distribution. Eq. (12a) indicates that the time-varying conditional mean of extremes depends on the location \( \mu_{it} \), scale \( \sigma_{it} \), and shape \( \xi_{it} \) parameters of the GEV distribution. Eq. (12b) shows that the time-varying conditional standard deviation (or volatility) of extremes is determined by the scale and shape parameters of the GEV:

\[
\text{Mean}_{\text{GEV}} = \mu_{it} + \frac{\sigma_{it}}{\xi_{it}} \left( \Gamma(1 - \xi_{it}) - 1 \right),
\]

\[
\text{Volatility}_{\text{GEV}} = \frac{\sigma_{it}}{\xi_{it}} \left\{ \Gamma(1 - 2\xi_{it}) - \left[ \Gamma(1 - \xi_{it}) \right]^2 \right\}^{0.5}.
\]

As mentioned in Appendix A, the mean and volatility of extremes in Eqs. (12a)–(12b) are based on the assumption that the scale and shape parameters are positive. As will be discussed in Section 4, the time-varying scale and shape parameter estimates from the ‘unrestricted’ Log-L function are found to be positive through time, indicating the reliability of this assumption.

To generate a time-varying conditional EVT, we first specify the location parameter of the GEV distribution as a linear function of the last period’s extreme returns. The first-order serial correlation between the extreme returns is modeled by introducing an autoregressive of order one, AR(1), process in \( \mu_{it} \):

\[
\mu_{it} = \mu + \phi M_{i,t-1}.
\]

Using Eq. (13), one can test whether the last period’s extremes comprise some significant information which can be used to explain the dynamic behavior of the current extremes. Specifically, one can test whether the coefficient \( \phi \), on \( M_{i,t-1} \), is statistically significant.

Second, we parameterize the current scale parameter of the GEV distribution, \( \sigma_{it} \), as a function of the last period’s unexpected news, \( e_{i,t-1} \), and the last period’s

---

11We should note that the time-variation in extreme value distributions is originally considered by Bali and Neftci (2003) for the generalized Pareto distribution (GPD) of Pickands (1975). They analyze the dynamic behavior of the maximal and minimal changes in daily federal funds rates, and find a significant time-series variation in the conditional mean and standard deviation of the excesses over high thresholds.

12Note that this paper is the first to derive the conditional mean and volatility of extremes based on the GEV distribution.
The scale parameter, $\sigma_{t-1}$:

$$\sigma_t = \sigma + \lambda_1 \epsilon_{t-1} + \lambda_2 \sigma_{t-1},$$  \hspace{1cm} (14)

where $\epsilon_t = M_{it} - \mu_{it}$ measures the deviation of the predicted maxima $\mu_{it}$ from the actual $M_{it}$, and can thus be viewed as an unexpected information shock to the stock market during its largest falls and rises. Eq. (14) models the scale parameter as a moving average of the past absolute shocks $\{|\epsilon_{t-1}|, |\epsilon_{t-2}|, \ldots\}$ since $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_1 + \lambda_2 < 1$. Using Eq. (14), one can determine the presence of volatility persistence in the standard deviation of extremes. Specifically, one can test the hypothesis $\lambda_1 = \lambda_2 = 0$ to see if the standard deviation of extremes is constant or accommodates volatility clustering.

Finally, we model the current shape parameter of the GEV distribution, $\xi_t$, as a function of the last period’s information shocks, $\epsilon_{t-1}$, and the last period’s shape parameter, $\xi_{t-1}$:

$$\xi_t = \xi + \gamma_1 |\epsilon_{t-1}| + \gamma_2 \xi_{t-1}.$$ \hspace{1cm} (15)

In Eq. (15), the tail index that measures the fatness of the distribution (or the weight of the tails) is parameterized as a moving average of the past absolute shocks $\{|\epsilon_{t-1}|, |\epsilon_{t-2}|, \ldots\}$. Based on the estimated values of $\gamma_1$ and $\gamma_2$, one can test whether the current weights of the tails are affected by the last period’s unexpected news and the last period’s tail-thickness parameter.

Two parametric approaches are commonly used to estimate the extreme value distributions: (1) the maximum likelihood method which yields parameter estimators which are unbiased, asymptotically normal, and of minimum variance, and (2) the regression method which provides a graphical method for determining the type of asymptotic distribution. In this paper the maximum likelihood method is used to estimate the conditional distribution of extreme high-frequency returns.

The GEV distribution in Eq. (11) has a conditional density function for the extremes

$$h(M_{it}; \mu_{it}, \sigma_{it}, \xi_{it}) = \left(\frac{1}{\sigma_{it}}\right) \left[1 + \xi_{it} \left(\frac{M_{it} - \mu_{it}}{\sigma_{it}}\right)\right]^{\left((1+\xi_{it})/\xi_{it}\right)} \times \exp\left\{ - \left[1 + \xi_{it} \left(\frac{M_{it} - \mu_{it}}{\sigma_{it}}\right)\right]^{-1/\xi_{it}} \right\},$$ \hspace{1cm} (16)

which gives the conditional Log-L function:

$$\log L_{GEV} = -T \ln \sigma_{it} - \left(\frac{1 + \xi_{it}}{\xi_{it}}\right) \sum_{i=1}^{T} \ln \left[1 + \xi_{it} \left(\frac{M_{it} - \mu_{it}}{\sigma_{it}}\right)\right]$$

$$- \sum_{i=1}^{T} \left[1 + \xi_{it} \left(\frac{M_{it} - \mu_{it}}{\sigma_{it}}\right)\right]^{-1/\xi_{it}}.$$ \hspace{1cm} (17)

Maximizing the conditional Log-L function in Eq. (17) with respect to the parameters $\mu$, $\phi$, $\sigma$, $\lambda_1$, $\lambda_2$, $\xi$, $\gamma_1$, and $\gamma_2$ yields time-varying scale ($\sigma_{it}$) and shape ($\xi_{it}$) parameters. Then, we substitute estimated $\sigma_{it}$ and $\xi_{it}$ in Eq. (12b) to obtain the
conditional EVT.\textsuperscript{13,14} When estimating the time-varying location, scale, and shape parameters in Eqs. (13)–(15), we do not impose any restrictions on the Log-L function or on the parameters. As will be discussed later, all the parameters from the unrestricted MLE methodology turn out to be positive, except for $\mu_{\text{min}} < 0$.

3. Data

There are three main types of data: daily index returns, daily implied volatilities, and intraday index returns. The first data set consists of daily closing levels of the S&P 100 index. The time period of investigation for daily index returns extends from 1/2/1987 to 8/31/2000, giving a total of 3426 observations.\textsuperscript{15} This data set is used to estimate the conditional volatilities based on the discrete time GARCH models with the generalized error, Student-$t$, and normal distributions. Because of the autoregressive of order one process in the GARCH (or TS-GARCH) models [see Eqs. (1)–(3)], 3426 daily index returns generate a time-series of 3425 daily volatilities from 1/5/1987–8/31/2000.

The second data set includes the CBOE’s, VIX, from 1/5/1987 to 8/31/2000, yielding a total of 3425 daily observations. Daily implied volatilities are computed from an annualized implied volatility index as $VIX = \sqrt{252}$.\textsuperscript{16}

The third data set contains high-frequency intraday data from 1/2/1987 to 8/31/2000, giving a total of 269,731 5-min returns. These 5-min returns are constructed from index levels recorded every 15 s. As discussed earlier, we use 5-min returns from the S&P 100 index to calculate a measure of realized (or integrated) volatility because this is the highest frequency that previous research uses. To construct the daily realized volatility, we square and then sum 5-min returns for the period from 9:30 EST to 16:00 EST. The original intraday data are also used to obtain the 1-day maxima and minima. To check the robustness of our findings on time-series variation in extreme value distributions of high-frequency returns, we use 5, 15, and 30-min returns. 3426 daily maxima and minima from 1/2/ 1987 to 8/31/2000 are utilized to estimate the conditional EVT.

Table 1 shows descriptive statistics for 5-min and daily returns. The unconditional mean of 5-min returns is about 0.0007% with a standard deviation of 0.10%. The unconditional mean of daily returns is about 0.057% with a standard deviation of 1.13%. The maximum and minimum values are about 2.25% and $-2.25\%$ for 5-min,

\textsuperscript{13}Differentiating the log-likelihood function in Eq. (17) with respect to the location, scale, and shape parameters yields the first-order conditions of the maximization problem. Clearly, no explicit solution exists to these nonlinear equations, and thus numerical procedures or search algorithms are called for. Details and presentation of alternative statistical estimation methods can be found in Leadbetter et al. (1983), Resnick (1987), Castillo (1988), and Embrechts et al. (1997).

\textsuperscript{14}The parameters are estimated using the Berndt et al. (1974) numerical algorithm and the standard errors are computed based on the outer-product of the gradients.

\textsuperscript{15}The October 1987 crash period is included in our analysis. To check the robustness of our results on the in-sample and out-of-sample performance of alternative volatility models, we repeated the analysis, excluding the October 1987 episode. The empirical findings turn out to be very similar to those reported in our tables. They are available upon request.
and about 8.53% and −23.78% for daily returns. The skewness and excess kurtosis statistics are reported for testing the distributional assumption of normality. The skewness statistics for 5-min and daily returns are close to zero, but significant at the 1% level. The excess kurtosis statistics are considerably high and significant at the 1% level, implying that the distribution of equity returns has much thicker tails than the normal distribution. The fat-tail property is more dominant than skewness in the sample. The Augmented Dickey-Fuller (ADF) statistic indicates strong rejection of the null hypothesis of a unit root for 5-min and daily returns. The autocorrelations of return series are generally small and not systematically positive or negative. The autocorrelations of volatility series are all positive and significant at the 1% level.

Table 1 also provides descriptive statistics for daily realized, implied, GED TS-GARCH, and EVT. A notable point is that the means, standard deviations, skewness, kurtosis, maximum, and minimum values of daily realized and EVT are very similar especially as compared to VIX and GED TS-GARCH. The kurtosis values for VIX and GED TS-GARCH models turn out to be considerably higher than those for the realized volatility. Overall, the skewness and kurtosis statistics of different volatility measures imply that the volatility distribution is skewed to the right and has much thicker tails than the normal distribution.
4. Estimation results

Table 2 presents the maximum likelihood estimates of the time-varying location ($\mu_t$), scale ($\sigma_t$), and shape ($\xi_t$) parameters of the GEV distribution. Asymptotic $t$-statistics are given in parentheses. The maximized Log-L values are reported for each model to test the presence of time-series variation in the location, scale and shape parameters. The GEV parameters are estimated using the 1-day maximal and minimal returns on the S&P 100 index.

Table 2
Maximum likelihood estimates of the GEV distribution

<table>
<thead>
<tr>
<th>Maxima</th>
<th>$\mu$</th>
<th>$\phi$</th>
<th>$\sigma$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\xi$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>Log-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_t, \sigma_t, \xi_t$</td>
<td>0.00148</td>
<td>0.2798</td>
<td>0.000018</td>
<td>0.0325</td>
<td>0.9263</td>
<td>0.0049</td>
<td>0.0453</td>
<td>0.8724</td>
<td>18726.68</td>
</tr>
<tr>
<td>(41.568)</td>
<td>(20.938)</td>
<td>(4.2332)</td>
<td>(5.4829)</td>
<td>(82.655)</td>
<td>(2.3106)</td>
<td>(2.4779)</td>
<td>(28.339)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_t, \sigma_t, \xi_t$</td>
<td>0.00146</td>
<td>0.2834</td>
<td>0.000024</td>
<td>0.0525</td>
<td>0.8873</td>
<td>0.1331</td>
<td>0.0</td>
<td>0.0</td>
<td>18712.44</td>
</tr>
<tr>
<td>(41.148)</td>
<td>(20.918)</td>
<td>(5.5538)</td>
<td>(8.4675)</td>
<td>(73.263)</td>
<td>(8.9569)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_t, \sigma_t, \xi_t$</td>
<td>0.00151</td>
<td>0.2982</td>
<td>0.000064</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2050</td>
<td>0.0</td>
<td>0.0</td>
<td>18540.12</td>
</tr>
<tr>
<td>(44.495)</td>
<td>(29.113)</td>
<td>(36.557)</td>
<td>(12.692)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minima</td>
<td>$\mu_t, \sigma_t, \xi_t$</td>
<td>-0.00137</td>
<td>0.3044</td>
<td>0.000053</td>
<td>0.0433</td>
<td>0.8527</td>
<td>0.0021</td>
<td>0.0548</td>
<td>0.9350</td>
</tr>
<tr>
<td>(-39.477)</td>
<td>(23.389)</td>
<td>(3.5818)</td>
<td>(5.1826)</td>
<td>(28.951)</td>
<td>(2.2875)</td>
<td>(2.4779)</td>
<td>(45.128)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_t, \sigma_t, \xi_t$</td>
<td>-0.00134</td>
<td>0.3133</td>
<td>0.000045</td>
<td>0.0501</td>
<td>0.8937</td>
<td>0.0907</td>
<td>0.0513</td>
<td>0.8428</td>
<td>18565.21</td>
</tr>
<tr>
<td>$\mu_t, \sigma_t, \xi_t$</td>
<td>-0.00140</td>
<td>0.3159</td>
<td>0.00065</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2224</td>
<td>0.0</td>
<td>0.0</td>
<td>18311.81</td>
</tr>
<tr>
<td>(-41.682)</td>
<td>(30.764)</td>
<td>(36.699)</td>
<td>(13.259)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table presents the maximum likelihood estimates of the time-varying location, scale, and shape parameters of the GEV distribution. The intraday (5-min) returns data cover the period from January 2, 1987 through August 31, 2000. The 1-day maximal and minimal returns on S&P 100 are obtained from 5-min returns observed in a day. 3426 daily extremes are used in maximum likelihood estimation. Asymptotic $t$-statistics are given in parentheses. The maximized likelihood values (Log-L) are reported for each model.

GEV distribution: $H(M_t; \mu_t, \sigma_t, \xi_t) = \exp \left\{ - \left[ 1 + \xi_t \left( \frac{M_t - \mu_t}{\sigma_t} \right) \right]^{-1/\xi_t} \right\}$

Location: $\mu_t = \mu + \phi M_{t-1}$, $\mu_t = \mu$ when $\phi = 0$. Scale: $\sigma_t = \sigma + \lambda_1 |\xi_{t-1}| + \lambda_2 \xi_{t-1}$, $\sigma_t = \sigma$ when $\lambda_1 = \lambda_2 = 0$. Shape: $\xi_t = \xi + \gamma_1 |\xi_{t-1}| + \gamma_2 \xi_{t-1}$, $\xi_t = \xi$ when $\gamma_1 = \gamma_2 = 0$. 

According to the asymptotic \( t \)-statistics shown in Table 2, all of the estimated GEV parameters are statistically significant at the 1\% or 5\% level. For the extreme return process with constant location \((\mu_{it} = \mu)\) and scale \((\sigma_{it} = \sigma)\) parameters, the estimates of \( \xi \) are positive and highly significant. Specifically, the estimated shape parameter for the maxima \((\xi_{\text{max}} = 0.22)\) is slightly greater than that for the minima \((\xi_{\text{min}} = 0.18)\).\(^{16}\) Since the higher \( \xi \), the fatter the distribution of extremes, the maximal returns have slightly thicker tails than the minimal returns.

A notable point in Table 2 is that for all specifications of the scale and shape parameters, the \( AR(1) \) coefficient, \( \phi \), in the time-varying location parameter \((\mu_{it})\) is statistically significant at the 1\% level for both the maximal and minimal returns. The maximized Log-L values of the models with constant \( \mu \) and time-varying \( \mu_{it} \) indicate strong rejection of the null hypothesis \( \phi = 0 \). The results imply the presence of first-order serial correlation in daily extremes, i.e., the last period’s maxima (or minima) comprise statistically significant information, which can be used to explain the dynamic behavior of the current extremes. Eq. (12a) shows that the first moment of the GEV distribution largely depends on the time-varying location parameter. Therefore, we can conclude that the conditional mean of the current maxima (or minima) depends on the last period’s extremes.

As presented in Table 2, the parameters \((\lambda_1, \lambda_2)\) are estimated to be positive, and highly significant, indicating substantial time-series variation in the scale parameter \((\sigma_{it})\). The maximized Log-L values of the models with constant \( \sigma \) and time-varying \( \sigma_{it} \) indicate strong rejection of the null hypothesis \( \lambda_1 = \lambda_2 = 0 \). A notable point in Table 2 is that the current scale parameter \((\sigma_{it})\) of the maxima and minima turns out to be more sensitive to the last period’s scale parameter \((\sigma_{it-1})\) than to the last period’s unexpected news \((e_{it-1})\) since \( \lambda_2 > \lambda_1 \) in all cases. Another notable point is that when the current scale parameter is defined as a function of \( e_{it-1} \) and \( \sigma_{it-1} \) as in Eq. (14), the constant shape parameter \((\xi)\) is estimated to be somewhat lower than in the model with constant scale parameter \( \sigma \). In other words, the tail-thickness of the maximal and minimal returns is reduced when conditional heteroscedasticity in extremes is modeled with the GARCH-type specification of the scale parameter.\(^{17}\) Table 2 also indicates that unexpected information shocks persist longer in the maxima than in the minima because the sum of the parameters \((\lambda_1 + \lambda_2)\) in \( \sigma_{it} \) is greater for the maximal returns than for the minimal returns. For all specifications considered in the paper, the sum \((\lambda_1 + \lambda_2)\) is found to be in the range of 0.94–0.96 for the maxima, and 0.89–0.93 for the minima.

Table 2 provides evidence that the current shape parameter \((\xi_{it})\) is highly affected by the last period’s unexpected news \((e_{it-1})\) and the last period’s shape parameter \((\xi_{it-1})\). The parameters \((\gamma_1, \gamma_2)\) are estimated to be positive and highly significant, implying considerable time-variation in the tail index. The maximized Log-L values

\(^{16}\)In all cases, the tail index \( \xi \) is estimated to be positive and statistically different from zero. Although not presented in the paper, the likelihood ratio (LR) test between the Frechet case and Gumbel case leads to a firm rejection of the Gumbel distribution (and a fortiori a rejection of the Weibull distribution).

\(^{17}\)The observed leptokurtosis in the maximal and minimal returns is reduced when the extremes are normalized by the time-varying scale parameter, which implies that the presence of excess kurtosis in extreme returns could be related to the time-variation in conditional volatility that depends on \( \mu \) and \( \mu_t \).
of the models with constant $\xi$ and time-varying $\xi_i$ indicate strong rejection of the null hypothesis $\gamma_1 = \gamma_2 = 0$. Similar to our findings for $\sigma_i$, the current tail-thickness parameter of the maxima and minima turns out to be more sensitive to the last period’s tail index ($\xi_{i-1}$) than to the last period’s unexpected news ($\epsilon_{i-1}$) since $\gamma_2 > \gamma_1$ without any exception. Another notable point in Table 2 is that when $\xi_i$ is specified as a function of $\epsilon_{i-1}$ and $\xi_{i-1}$ as in Eq. (15), the degree of first-order serial correlation in extremes is slightly reduced. Specifically, the magnitude of the $AR(1)$ coefficient, $\phi$, in $\mu_i = \mu + \phi M_{t-1}$ turns out to be somewhat lower than in the model with constant $\sigma$ and $\xi$. Table 2 also shows that unexpected information shocks persist longer in the tails of the minima than in the tails of the maxima because the sum of the parameters ($\gamma_1 + \gamma_2$) in $\xi_i$ is greater for the minimal returns than for the maximal returns. For all specifications considered in the paper, the sum ($\gamma_1 + \gamma_2$) is in the range of 0.98–0.99 for the minima, and 0.89–0.92 for the maxima.

These results point to the presence of volatility clustering and persistence (i.e., ARCH effects) in the standard deviation of high-frequency extreme returns since the conditional volatility of extremes is determined by $\sigma_i$ and $\xi_i$ that are characterized by substantial persistence of past information shocks.

Figs. 1A, B and C display the time-varying location, scale, and shape parameters of the GEV distribution, respectively. These parameters are estimated without any restrictions on $\mu$, $\phi$, $\sigma$, $\lambda_1$, $\lambda_2$, $\xi$, $\gamma_1$, and $\gamma_2$ given in Eqs. (13), (14), and (15). As discussed earlier, the parameters $\sigma$, $\lambda_1$, and $\lambda_2$ in Eq. (14) should be positive for the time-varying scale parameter to be positive: $\sigma_i > 0$. Similarly, the parameters $\xi$, $\gamma_1$, and $\gamma_2$ in Eq. (15) should be positive for the time-varying shape parameter to be positive: $\xi_i > 0$. As shown in Figs. 1B and C, both $\sigma_i$ and $\xi_i$ from the unrestricted MLE methodology turn out to be positive through time without any exception. The time-varying location parameter $\mu_i$ is also estimated without any restrictions and as shown in Fig. 1A, $\mu_i$ is positive for the maxima and negative for the minima through time. These results provide supporting empirical evidence for the assumption that the scale and shape parameters be positive.

5. Performance of alternative volatility models

5.1. In-sample

The in-sample performance of implied, GARCH, and EVT$s is first measured by testing their predictive power for the square-root of the sum of squared 5-min returns on the S&P 100 index. The conditional standard deviations of the maximal and minimal returns based on the time-varying location, scale, and shape parameters turn out to be very similar. However, the left tail of the empirical distribution (i.e., the minimal returns) is more extensively studied by the former researchers (especially

\[ 18^{18} \text{To save space, we do not present the maximum likelihood estimates of the symmetric and asymmetric GARCH models with GED, Student-} \tau, \text{and Normal density. They are available upon request.} \]
for financial risk management calculations). Therefore, we prefer to present the relative performance of EVT based on the minimal returns.

We compute the proportion of the total variation in daily realized volatility that can be explained by the estimated conditional variances (or standard deviations), and report the adjusted-$R^2$ values. The predictive power of alternative models is

![Graph A](image1)

**Fig. 1.** (A) Time-varying location parameter for maxima and minima. (B) Time-varying scale parameter for maxima and minima. (C) Time-varying shape parameter for maxima and minima.
evaluated by estimating a series of OLS regressions of the form:

\[
\text{Realized}_t = \omega_0 + \omega_1 \text{VIX}_t + \omega_2 \text{EVT}_{t-1} + \omega_3 \text{GARCH}_{t-1} + u_t,
\]

(18)

where

\[
\text{Realized}_t = \sqrt{\sum_{q=1}^{79} \frac{R^2_{(q/79),t}}{q}}
\]

is the square-root of the sum of squared 5-min returns on S&P 100 index at time \( t \). \( \text{VIX}_t \) is the daily VIX at time \( t \), \( \text{EVT}_{t-1} \) is the daily EVT at time \( t \) given the information set until time \( t-1 \), \( \text{GARCH}_{t-1} \) is the daily GARCH volatility of index returns at time \( t \) given the information set until time \( t-1 \), and \( u_t \) is the forecast error.\(^{19}\)

Since we predict the same dependent variable (i.e., realized volatility) with different sets of regressors (i.e., different combinations of \( \text{VIX}_t \), \( \text{EVT}_{t-1} \), and \( \text{GARCH}_{t-1} \)), and evaluate the relative performance of these volatility estimators, we choose to use the adjusted-\( R^2 \) (instead of the unadjusted-\( R^2 \)). The adjusted-\( R^2 \) takes explicit account of the number of regressors used in the equation. It is useful for comparing the fit of specifications that differ in the addition or deletion of

\(^{19}\)At an earlier stage of the study we find that the GED TS-GARCH model outperforms alternative GARCH models with Student-\( t \) or Normal density in forecasting realized volatility. Therefore, in Eq. (18), we prefer to use the GED TS-GARCH model for \( \text{GARCH}_{t-1} \) in evaluating the relative performance of implied, GARCH, and extreme value volatility estimators.
explanatory variables. The unadjusted-$R^2$ will never decrease with the addition of any variable to the set of regressors. If the added variable is totally irrelevant, the explained sum of squares simply remains constant. The adjusted-$R^2$, however, may decrease with the addition of variables of low explanatory power. Hence, we use the adjusted-$R^2$ values to compare the real contribution of each volatility estimator to the prediction of realized volatility.

In addition to comparing the adjusted-$R^2$ values of the regressions for each volatility model, we test the statistical significance of $\omega_1$, $\omega_2$, and $\omega_3$. First, we estimate the univariate regressions of the form, $\text{Realized}_t = \omega_0 + \omega_1 VIX_t + u_t$, $\text{Realized}_t = \omega_0 + \omega_2 EVT_{t|t-1} + u_t$, $\text{Realized}_t = \omega_0 + \omega_3 GARCH_{t|t-1} + u_t$ and then determine the statistical significance of $\omega_1$, $\omega_2$, and $\omega_3$ using the standard $t$-test. Standard errors of the regression coefficients are calculated using the procedure of Newey and West (1987). As shown in Table 3, VIX performs better than GARCH and EVT estimators based on the adjusted determination coefficients. The adjusted-$R^2$ values for in-sample forecasts are about 48.47% for VIX, 41.74% for EVT, and 40.17% for the GARCH model. In univariate regressions, the coefficients ($\omega_1$, $\omega_2$, $\omega_3$) on $VIX_t$, $EVT_{t|t-1}$, and $GARCH_{t|t-1}$ are found to be significant at the 1% level.

Second, the two-variable regressions of the form: $\text{Realized}_t = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_{t|t-1} + u_t$ and $\text{Realized}_t = \omega_0 + \omega_1 VIX_t + \omega_3 GARCH_{t|t-1} + u_t$ are estimated

<table>
<thead>
<tr>
<th>In-sample forecast</th>
<th>$\omega_0$</th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$VIX_t$</td>
<td>0.00041</td>
<td>0.5551</td>
<td>0.0</td>
<td>0.0</td>
<td>0.484690</td>
<td>28.20%</td>
<td>42.52%</td>
<td>45.35%</td>
</tr>
<tr>
<td></td>
<td>(0.8753)</td>
<td>(13.978)</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$EVT_t$</td>
<td>0.00104</td>
<td>0.0</td>
<td>0.6673</td>
<td>0.0</td>
<td>0.417374</td>
<td>19.21%</td>
<td>28.36%</td>
<td>30.90%</td>
</tr>
<tr>
<td></td>
<td>(2.4930)**</td>
<td>(14.338)**</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$GARCH_t$</td>
<td>0.00218</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5332</td>
<td>0.401723</td>
<td>20.67%</td>
<td>30.67%</td>
<td>35.14%</td>
</tr>
<tr>
<td></td>
<td>(0.3089)**</td>
<td>(7.1531)**</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t$, $EVT_t$</td>
<td>-0.00029</td>
<td>0.4007</td>
<td>0.2747</td>
<td>0.0</td>
<td>0.515158</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(-0.6283)</td>
<td>(10.279)**</td>
<td>(7.0110)**</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t$, $GARCH_t$</td>
<td>0.00051</td>
<td>0.4646</td>
<td>0.0</td>
<td>0.1052</td>
<td>0.487925</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(1.0595)</td>
<td>(9.2126)**</td>
<td>(1.2938)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t$, $EVT_t$,</td>
<td>-0.00033</td>
<td>0.4168</td>
<td>0.2830</td>
<td>-0.0242</td>
<td>0.515167</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$GARCH_t$</td>
<td>(-0.8839)</td>
<td>(10.089)**</td>
<td>(5.9699)**</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table shows the in-sample performance of alternative volatility models. The predictive power of alternative models is evaluated by estimating an OLS regression: $\text{Realized}_t = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_{t|t-1} + \omega_3 GARCH_{t|t-1} + u_t$, $\text{Realized}_t$ is the square-root of the sum of squared 5-min returns on S&P 100 index at time $t$, $VIX_t$ is the daily implied volatility index at time $t$, $EVT_{t|t-1}$ is the daily extreme value volatility estimator at time $t$, $GARCH_{t|t-1}$ is the daily volatility of index returns at time $t$, and $u_t$ is the forecast error. Adjusted-$R^2$ values indicate the proportion of the total variation in daily realized volatility that can be explained by $VIX_t$, $EVT_{t|t-1}$, and $GARCH_{t|t-1}$. The Newey–West (1987) adjusted $t$-statistics of the regression coefficients are given in parentheses. *, **, *** denote statistical significance at least at the 10% (5%) [1%] level. Theil Inequality Coefficient (TIC) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) measure how far the estimated volatilities are away from the realized (or integrated) volatility.
to determine the marginal contribution of $EVT_{t-1}$ and $GARCH_{t-1}$. Table 3 shows the adjusted-$R^2$ increased only by 0.37% (from 48.47% to 48.79%) after including $GARCH_{t-1}$ to the univariate regression equation $Realized_t = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_{t-1} + \epsilon_t$. In addition, the coefficient $\omega_3$ on $GARCH_{t-1}$ is found to be statistically insignificant based on the Newey–West t-statistic = 1.29, and the Wald statistic = 1.67 with a p-value of 0.1957. Table 3 indicates statistically significant contribution of EVT: adjusted-$R^2$ increased by 3.05% (from 48.47% to 51.52%), and the coefficient $\omega_2$ on $EVT_{t-1}$ is found to be highly significant based on the Newey–West t-statistic = 7.01, and the Wald statistic = 49.15 with a p-value of zero. The results indicate a strong rejection of the null hypothesis $\omega_2 = 0$, and a failure to reject the null $\omega_3 = 0$ even at the 10% level.

Finally, we estimate the regression Eq. (16) and test the statistical significance of $\omega_2$ and $\omega_3$ using the standard $t$ and Wald tests. Table 3 implies statistically significant contribution of $EVT_{t-1}$, and almost no contribution of $GARCH_{t-1}$ to the explanation of realized volatility. A notable point is that the adjusted-$R^2$ = 51.52% remains almost the same after including $GARCH_{t-1}$ to the bivariate regression equation $Realized_t = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_{t-1} + \epsilon_t$, and the coefficient $\omega_3$ on $GARCH_{t-1}$ is found to be statistically insignificant based on the Newey–West t-statistic = −0.33 and the Wald statistic = 0.11 with a p-value of 0.7430. Table 3 implies strong rejection of the null hypothesis $\omega_2 = 0$ because the coefficient $\omega_2$ on $EVT_{t-1}$ remains highly significant based on the Newey–West t-statistic = 5.97 and the Wald statistic = 35.64 with a p-value of zero.

As discussed in Andersen et al. (2005), the integrated volatility is measured with error and the exact form of the measurement error depends on the assumed model structure (see, e.g., Meddahi (2002) and Barndorff-Nielsen and Shephard (2002a)). Andersen et al. (2005) indicate that if the realized volatility is used in place of the true (latent) integrated volatility, the usual $R^2$ values will be downward biased. They focus their forecast comparisons on the $R^2$ in the Mincer-Zarnowitz (1969) style regressions of the ex-post realized volatility on the corresponding model forecasts. They show that the (feasible) $R^2$ from the commonly used Mincer–Zarnowitz regression will underestimate the true predictability as measured by the (infeasible) $R^2$ from the regression of the future (latent) integrated volatility on the same set of volatility forecasts by the ‘multiplicative factor’:

$$\frac{Var[Realized_t(h)]}{Var[Realized_t(h)] - 2hE[RQ_t(h)],}$$

where $Realized_t(h) \equiv \sum_{i=1}^{1/h} R_{t-1+h}^{(h)}$ is the realized variance defined by the summation of the $1/h$ intra-period squared returns (where $h = 1/q$), $R_t^{(h)} \equiv \ln(P_t) - \ln(P_{t-h})$, and $RQ_t(h) \equiv 1/3h \sum_{i=1}^{1/h} R_{t-1+h}^{(h)}$ is the realized quarticity. For the 5-min returns on the S&P 100 index from 1/5/1987 to 8/31/2000, we calculated the multiplicative factor:
and find that the corrected $R^2$ values exceed their unadjusted counterparts by 7.2%.

At an earlier stage of the study, we compute the unadjusted-$R^2$ values corresponding to the adjusted-$R^2$'s given in Table 3. We do not report the unadjusted-$R^2$ values because the difference between the adjusted and unadjusted $R^2$ values are very small, in the range of 0.01–0.04%. This is simply because the number of explanatory variables in Eq. (18) and its restricted versions is very small compared to the number of observations. Thus, one can multiply the adjusted-$R^2$ values shown in Table 3 by 1.072 to compute the corrected $R^2$ values of Andersen et al. (2005).

Further useful insights can be gained from Figs. 2A–C, which present the time-series plots of realized volatility, VIX, and the estimated conditional volatilities of EVT and GED TS-GARCH models. The striking observation in Fig. 2A is that even though VIX provides the most accurate in-sample forecasts based on the adjusted-$R^2$ values, it is still a biased predictor of realized volatility. As shown in Fig. 2A, $VIX_t$ overestimates realized volatility, $Realized_t$, in most periods from 1/5/1987 to 8/31/2000. The average value of $VIX_t/Realized_t$ is about 1.87 for the sample period. In contrast to implied volatility in Fig. 2A, the EVT in Fig. 2B and the GED TS-GARCH model in Fig. 2C track realized volatility much better.

Figs. 2A–C imply that if we remain within the statistical tradition of reporting summary statistics based directly on the deviation between forecasts and realizations, the relative performance of alternative volatility models may change. Therefore, as an alternative to $R^2$ measures, we compute the Theil’s (1961) Inequality Coefficient (TIC), which always lies between zero and one, where zero indicates a perfect fit.

$$TIC = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (vol_{realized,t} - vol_{estimated,t})^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (vol_{realized,t})^2} + \sqrt{\frac{1}{n} \sum_{t=1}^{n} (vol_{estimated,t})^2}},$$

(20)

where $vol_{realized,t}$ is the realized volatility of daily returns at time $t$, $vol_{estimated,t}$ is the estimated conditional variance (or standard deviations) at time $t$ given the information set until time $t-1$ based on the implied, EVT, or GARCH models, i.e., $vol_{estimated,t} = VIX_t, EVT_{t-1},$ or $GARCH_{t-1}$.

The results in Table 3 highlight the superior in-sample performance of EVT against the implied (VIX) and GARCH volatility models. Specifically, the TIC values for in-sample forecasts of realized volatility are found to be 19.21% for EVT, 28.20% for VIX, and 20.67% for the GED TS-GARCH model. As expected from Figs. 2A–C, VIX, turns out to be inferior to EVT and GARCH based on the deviation between forecasts and realizations.

---

20Theil’s inequality coefficient given in Eq. (20) is originally proposed by Theil (1961), and then used by Stekler (1968), Bliemel (1973), Leuthold (1975), and Pindyck and Rubinfeld (1991) among many others.
As an alternative to TIC, we use the heteroscedasticity-adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) to measure how far the estimated volatilities are away from the realized volatility. In other words, we compute the deviation between $\text{vol}_{\text{realized}, t}$ and $\text{vol}_{\text{estimated}, t}$ using the mean absolute percentage errors (HMAE) and the root mean squared percentage

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{(A) Realized volatility versus implied volatility (VIX). (B) Realized volatility versus EVT. (C) Realized volatility versus GED-TS-GARCH.}
\end{figure}
errors (HRMSE)\textsuperscript{21}:

\[ HMAE = \frac{1}{n} \sum_{t=1}^{n} \left| 1 - \frac{\text{vol}_{\text{realized},t}}{\text{vol}_{\text{estimated},t}} \right|, \] (21)

\[ HRMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left( 1 - \frac{\text{vol}_{\text{realized},t}}{\text{vol}_{\text{estimated},t}} \right)^2}. \] (22)

These two volatility forecast evaluation criteria follow the statistical tradition of reporting statistics based directly on the deviation between forecasts and realizations, while adjusting for heteroscedasticity in the forecast error. Table 3 provides the HMAE and HRMSE values for in-sample forecasts. The results are very similar to our earlier findings from TIC: the EVT performs better than the implied and GARCH volatility models, and the GED TS-GARCH model provides more accurate forecasts of realized volatility than VIX. More specifically, the HMAE values for in-sample forecasts of realized volatility are 28.36% for EVT, 42.52% for VIX, and 30.67% for the GARCH model. The HRMSE measures are 30.90% for EVT, 45.35% for VIX, and 35.14% for the GARCH model.

\textsuperscript{21}The HMAE and HRMSE are used by Andersen et al. (1999), and Andreou and Ghysels (2002).
5.2. Out-of-sample

The out-of-sample performance of alternative volatility models is evaluated based on their 1-day and 20-day-ahead forecasts of realized variance and standard deviation. The in-sample period is from 1/2/1987 to 12/31/1990 providing 1009 daily observations, followed by the out-of-sample period from 1/2/1991 to 8/31/2000, yielding 2417 1-day-ahead and 2398 20-day-ahead forecasts.\(^{22}\) To compare the out-of-sample forecasts of implied, GARCH, and EVTs, discrete time GARCH-in-mean models are estimated for daily index returns from 1/2/1987 to 12/31/1990. The most general specification of alternative models is as follows:

\[
R_t = \alpha_0 + \alpha_1 \sigma_{t/t-1} + \varepsilon_t \sigma_{t/t-1}. \tag{23}
\]

Conditional Variance:

\[
\sigma^2_{t/t-1} = \beta_0 + \beta_1 \varepsilon^2_{t-1} \sigma^2_{t-1} + \beta_2 \sigma^2_{t-1} + \delta VIX^2_{t-1} + \gamma EVT^2_{t-1}. \tag{24}
\]

Conditional Std. Dev.:

\[
\sigma_{t/t-1} = \beta_0 + \beta_1 \varepsilon_{t-1} \sigma_{t-1} + \beta_2 \sigma_{t-1} + \delta VIX_{t-1} + \gamma EVT_{t-1}, \tag{25}
\]

where \(\sigma^2_{t/t-1}\) and \(\sigma_{t/t-1}\) are the conditional variance and standard deviation of daily index returns based on the information set until time \(t-1\). In Eqs. (23)–(25), \(\varepsilon_t\) is drawn from the GED density. Note that imposing restrictions on certain parameters in Eqs. (23)–(25) three different conditional variance (or standard deviation) models are obtained using different daily information sets: (1) The GARCH(1,1) model of Bollerslev (1986) or TS-GARCH model of Taylor (1986) and Schwert (1989) with \(\delta = \gamma = 0\); (2) A volatility model that uses the latest information in the VIX series alone, \(\beta_1 = \beta_2 = \gamma = 0\); (3) A volatility model that uses information in daily extreme returns alone, \(\beta_1 = \beta_2 = \delta = 0\).

Time-series forecasts are obtained by estimating rolling VIX, EVT, GARCH, and TS-GARCH volatility models. Each conditional variance and standard deviation model is estimated initially over the 1009 trading days of the in-sample period from 1/2/1987 to 12/31/1990, and forecasts of realized variance and standard deviation are made for the next day, say \(t+1\), using the in-sample parameter estimates. The model and data are then rolled forward one day, deleting the observation(s) at time \(t-1008\) and adding on the observation(s) at time \(t+1\), reestimated and a forecast is made for time \(t+2\). This rolling method is repeated until the end of the out-of-sample forecast period. The 1-day-ahead forecasts provide predictions for 12/2/1991–8/31/2000 providing time-series of length 2,417. On each day, forecasts are also made for 20-day volatility.

Forecasts are produced at time \(t\) for the realized variance and standard deviation, defined by the sum of squared 5-min returns, \(\text{Variance}_t = \sum_{q=1}^{79} R^2_{(q/79),t}\) and

\[^{22}\text{Since the GARCH models generally need a considerably long sample to be estimated well, we use at an earlier stage of the study a longer in-sample period from 1/2/1987 to 12/31/1992 providing 1507 daily observations, followed by the out-of-sample period from 1/2/1993 to 8/31/2000, yielding 1919 1-day-ahead and 1900 20-day-ahead forecasts. The 1-day and 20-day-ahead forecasts of the GARCH models from the longer sample are found to be slightly more accurate than those reported in Table 4.}\]
For 20-day-ahead forecasts, realized variance and standard deviations are defined as
\[
Std Dev_t = \sqrt{\sum_{q=1}^{79} R_{(q/79),t}^2},
\]
where \( R_{(q/79),t} \) is the 5-min index return observed in day \( t \).
For 20-day-ahead forecasts, realized variance and standard deviations are defined as
\[
\text{Realized Var}_t \quad \text{and} \quad \text{Realized Std Dev}_t,
\]
respectively. Hence, our goal is to forecast the realized variance (and standard deviation) of 20-day holding period return.

The 1-day-ahead forecasts of GARCH and TS-GARCH models for \( \text{Variance}_t \), and \( \text{Std Dev}_t \) are defined by the recursive formulas
\[
\sigma_{t+1|t}^2 = \beta_0 + \beta_1 R_t^2 + \beta_2 \sigma_t^2 \quad \text{and} \quad \sigma_{t+1|t} = \beta_0 + \beta_1 |R_t| + \beta_2 \sigma_t,
\]
respectively. Forecasts for 20-day volatility are produced by aggregating expectations \( E(\sigma_{t+j|t}^2) = \beta_0 + \Theta E(\sigma_{t+j-1|t}^2) \) for variance and \( E(\sigma_{t+j|t}) = \beta_0 + \Theta E(\sigma_{t+j-1|t}) \) for standard deviation, where \( \Theta \) is the volatility persistence for GARCH(1,1) and TS-GARCH(1,1) models. As shown in Bali (2005), \( \Theta = \beta_2 + \beta_1 \frac{\Gamma(2/v)}{[\Gamma(3/v)]^{1/2}[\Gamma(1/v)]^{1/2}} \)
with GED density, and
\[
\Theta = \beta_2 + \beta_1 \left[ 2\Gamma\left(\frac{v+1}{2}\right) \Gamma\left(\frac{v}{2}\right)^{-1} \sqrt{\frac{v}{(v-1)^2\pi}} \right]
\]
with Student-\( t \) density.

For GARCH volatility forecasts of the realized variance over the coming 20 days, we sum the multi-step variance forecasts to obtain the volatility of the multi-period return. More specifically, we compute the 20-day-ahead conditional variance forecast of index returns by summing the one-day-ahead, two-day-ahead, etc., forecasts over the next 20 days and obtain a 20-day-ahead GARCH variance forecast. Once we have the 20-day-ahead GARCH variance forecast, we define the 20-day-ahead standard deviation estimator of the GARCH model as the square-root of the variance forecast.

The 1-day-ahead forecast of the volatility model that uses information contained in VIX alone is obtained by the recursive formula \( \sigma_{t+1|t}^2 = \beta_0 + \delta VIX_t^2 \) for realized variance, and by \( \sigma_{t+1|t} = \beta_0 + \delta VIX_t \) for realized standard deviation. Following Blair et al. (2001), we assume that \( E(\sigma_{t+j|t}^2) = E(\sigma_{t+j-1|t}^2) \) and \( E(\sigma_{t+j|t}) = E(\sigma_{t+j-1|t}) \) so that the 1-day-ahead forecast of VIX is multiplied by 20 to produce 20-day variance forecast and by \( \sqrt{20} \) to obtain 20-day standard deviation forecast. Similarly, the 1-day-ahead forecasts of EVTs that use information contained in the maximal and minimal returns are obtained by the recursive formula \( \sigma_{t+1|t}^2 = \beta_0 + \delta EVT_t^2 \) for realized variance and \( \sigma_{t+1|t} = \beta_0 + \delta EVT_t \) for realized standard deviation. To produce 20-day-ahead forecasts of realized volatility, we use the aforementioned multiplicative method.

We compute the proportion of the total variation in daily-realized variance (standard deviation) that can be explained by the estimated conditional variances...
The adjusted-$R^2$ values in this section provide information about how well each model is able to forecast the future volatility of stock market returns. The predictive power of alternative volatility models is evaluated by estimating a series of OLS regressions of the form:

$$Realized_{t+j} = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_t + \omega_3 GARCH_t + u_t,$$

where $j = 1$ or 20 days, $Realized_{t+j}$ represents realized standard deviation.

As presented in Table 4, the implied VIX performs better than the GARCH and EVT estimators based on the adjusted determination coefficients. More specifically, the adjusted-$R^2$ values for 1-day-ahead forecasts are about 56.69% for VIX, 49.87% for EVT, and 40.41% for the GARCH model. The two-variable regressions, $Realized_{t+j} = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_t + u_t$ and $Realized_{t+j} = \omega_0 + \omega_1 VIX_t + \omega_3 GARCH_t + u_t$, are estimated to determine the marginal contribution of $EVT_t$ and $GARCH_t$. Table 4 shows the adjusted-$R^2$ increased only by 0.06% (from 56.69% to 56.75%) after including $GARCH_t$ to the univariate regression equation $Realized_{t+j} = \omega_0 + \omega_1 VIX_t + u_t$. In addition, the coefficient $\omega_3$ on $GARCH_t$ is found to be statistically insignificant. Table 4 indicates statistically significant contribution of EVT: adjusted-$R^2$ increased from 56.69% to 58.65%, and the coefficient $\omega_2$ on $EVT_t$ is found to be highly significant. We also estimate the regression Eq. (26) and test the statistical significance of $\omega_2$ and $\omega_3$ using the standard $t$ and Wald tests. Table 4 implies statistically significant contribution of $EVT_t$, and almost no contribution of $GARCH_t$ to the explanation of realized volatility. A notable point is that the adjusted-$R^2 = 58.65\%$ remain almost the same after including $GARCH_t$ to the bivariate regression $Realized_{t+j} = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_t + u_t$, and the coefficient $\omega_3$ on $GARCH_t$ is found to be statistically insignificant.

Christensen and Prabhala (1998) indicate that the previous studies by Day and Lewis (1992) and Lamoureux and Lastrapes (1993) are characterized by a ‘maturity mismatch’ problem, in that Lamoureux and Lastrapes examine one-day-ahead and Day and Lewis examine 1-week-ahead predictive power of implied volatilities computed from options that have a much longer remaining life. In addition to 1-day-ahead forecasts, following Blair et al. (2001), we test the relative performance of alternative models for the longer forecast horizon of 20 trading days, that closely matches the life of the hypothetical option (22 trading days) that defines VIX. As presented in Table 4, the adjusted $R^2$ values and the related statistics for 20-day-ahead forecasts imply superior performance of VIX and the EVT in capturing time-series variation in realized volatility. The predictive power of discrete-time GARCH models with GED, Student-$t$, and normal distributions turns out to be inferior to VIX and EVT. In fact, there is only minor incremental information in GARCH models for 20-day-ahead forecasts, and this information is subsumed by implied and EVT.s.

As an alternative to $R^2$ measures, we compute the TIC for 1-day and 20-day-ahead forecasts of alternative models. The results in Table 4 highlight the superior

---

23To save space, we do not present results for the realized variance. They are available upon request.
Table 4
Out-of-sample performance of alternative volatility models

<table>
<thead>
<tr>
<th>1-day-ahead forecast</th>
<th>$\omega_0$</th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$VIX_t$</td>
<td>$-0.00143$</td>
<td>$0.7493$</td>
<td>$0.0$</td>
<td>$0.0$</td>
<td>$0.566899$</td>
<td>$24.83%$</td>
<td>$39.64%$</td>
<td>$42.54%$</td>
</tr>
<tr>
<td></td>
<td>($-3.7283$)**</td>
<td>($13.978$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$EVT_t$</td>
<td>$-0.00069$</td>
<td>$0.0$</td>
<td>$0.8662$</td>
<td>$0.0$</td>
<td>$0.498706$</td>
<td>$17.49%$</td>
<td>$27.31%$</td>
<td>$31.41%$</td>
</tr>
<tr>
<td></td>
<td>($-1.5446$)***</td>
<td></td>
<td>($16.381$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$GARCH_t$</td>
<td>$0.00110$</td>
<td>$0.0$</td>
<td>$0.0$</td>
<td>$0.6653$</td>
<td>$0.404105$</td>
<td>$19.54%$</td>
<td>$29.96%$</td>
<td>$34.42%$</td>
</tr>
<tr>
<td></td>
<td>($2.4735$)**</td>
<td></td>
<td></td>
<td>($13.388$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, EVT_t$</td>
<td>$-0.00174$</td>
<td>$0.6076$</td>
<td>$0.2084$</td>
<td>$0.0$</td>
<td>$0.586463$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>($-4.6896$)***</td>
<td>($18.394$)***</td>
<td>($5.4951$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, GARCH_t$</td>
<td>$-0.00152$</td>
<td>$0.7094$</td>
<td>$0.0$</td>
<td>$0.0594$</td>
<td>$0.567462$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>($-3.8818$)***</td>
<td>($11.389$)***</td>
<td></td>
<td>($0.8370$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, EVT_t, GARCH_t$</td>
<td>$-0.00172$</td>
<td>$0.6205$</td>
<td>$0.2181$</td>
<td>$-0.0289$</td>
<td>$0.586510$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>($-4.8345$)***</td>
<td>($12.515$)***</td>
<td>($4.4956$)***</td>
<td>($-0.4028$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20-day-ahead forecast</th>
<th>$\omega_0$</th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$VIX_t$</td>
<td>$-0.00037$</td>
<td>$0.6596$</td>
<td>$0.0$</td>
<td>$0.0$</td>
<td>$0.439539$</td>
<td>$32.53%$</td>
<td>$43.61%$</td>
<td>$47.80%$</td>
</tr>
<tr>
<td></td>
<td>($-0.9284$)***</td>
<td>($16.918$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$EVT_t$</td>
<td>$0.00049$</td>
<td>$0.0$</td>
<td>$0.7414$</td>
<td>$0.0$</td>
<td>$0.365477$</td>
<td>$24.44%$</td>
<td>$32.52%$</td>
<td>$39.20%$</td>
</tr>
<tr>
<td></td>
<td>($0.9125$)</td>
<td></td>
<td>($12.007$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$GARCH_t$</td>
<td>$0.00207$</td>
<td>$0.0$</td>
<td>$0.0$</td>
<td>$0.5640$</td>
<td>$0.289981$</td>
<td>$28.76%$</td>
<td>$34.55%$</td>
<td>$41.46%$</td>
</tr>
<tr>
<td></td>
<td>($4.4660$)***</td>
<td></td>
<td></td>
<td>($10.836$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, EVT_t$</td>
<td>$-0.00063$</td>
<td>$0.5420$</td>
<td>$0.1732$</td>
<td>$0.0$</td>
<td>$0.453195$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>($-1.5493$)***</td>
<td>($13.499$)***</td>
<td>($3.8397$)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, GARCH_t$</td>
<td>$-0.00054$</td>
<td>$0.5913$</td>
<td>$0.0$</td>
<td>$0.1039$</td>
<td>$0.440768$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>($-1.1917$)***</td>
<td>($10.209$)***</td>
<td></td>
<td>($1.3278$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$VIX_t, EVT_t, GARCH_t$</td>
<td>$-0.00068$</td>
<td>$0.5107$</td>
<td>$0.1523$</td>
<td>$0.0661$</td>
<td>$0.453209$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>($-1.6506$)***</td>
<td>($8.3707$)***</td>
<td>($2.6353$)***</td>
<td>($0.6744$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table shows the empirical performance of alternative volatility models based on their 1-day- and 20-day-ahead forecasts of realized volatility. The predictive power of alternative models is evaluated by estimating an OLS regression: $\text{Realized}_{t+j} = \omega_0 + \omega_1 VIX_t + \omega_2 EVT_t + \omega_3 GARCH_t + u_t$, where $j = 1$ or $20$, $\text{Realized}_{t+j}$ represents realized standard deviation, $VIX_t$ is the daily implied volatility index at time $t$, $EVT_t$ is the daily extreme value volatility estimator at time $t$, and $u_t$ is the forecast error. Adjusted-$R^2$ values indicate the proportion of the total variation in daily realized volatility that can be explained by $VIX_t$, $EVT_t$, and $GARCH_t$. It provides information about how well each model is able to forecast the future volatility of stock market returns. The Newey-West (1987) adjusted t-statistics of the regression coefficients are given in parentheses. * (**, ***) denote statistical significance at least at the 10% (5%) (1%) level. Theil Inequality Coefficient (TIC) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) measure how far the estimated volatilities are away from the realized (or integrated) volatility.
performance of EVT against the implied (VIX) and GARCH volatility models. Specifically, the TIC values for 1-day-ahead forecasts of realized standard deviation are found to be 17.49% for EVT, 24.83% for VIX, and 19.54% for the GARCH model. VIX turns out to be inferior to EVT and GARCH based on the deviation between forecasts and realizations.

We also use the HMAE and HRMSE to measure how far the estimated volatilities are away from the realized variance and standard deviation. Table 4 provides the HMAE and HRMSE values for 1-day and 20-day-ahead forecasts. The results are very similar to our earlier findings from TIC: the EVT perform better than the implied and GARCH volatility models, and the GED TS-GARCH model provides more accurate forecasts of realized volatility than VIX. More specifically, the HMAE values for 1-day-ahead forecasts of realized standard deviation are 27.31% for EVT, 39.64% for VIX, and 29.96% for the GARCH model. The HRMSE measures are 31.41% for EVT, 42.54% for VIX, and 34.42% for the GARCH model.

The TIC, HMAE, and HRMSE measures for 20-day-ahead forecasts indicate the same ranking of models obtained from 1-day-ahead forecasts. The results in Table 4 imply superior performance of EVT and GED TS-GARCH against VIX in capturing time-series variation in realized volatility.

6. Extensions and robustness

We extend the set of alternative volatility models to include forecasts from (i) modeling realized volatility directly and (ii) estimating the unconditional EVT estimator. We also consider the robustness of our results to (i) a different measure of realized volatility, and (ii) possible biases induced by the dependence structure of intraday returns.

6.1. Modeling and forecasting realized volatility directly

This section models and forecasts realized volatility directly. We use an ARMA\((p,q)\) process for realized volatility and compare the empirical performance of the ARMA\((p,q)\) realized volatility model with the implied, extreme value, and GED TS-GARCH volatility estimators. Specifically, the following ARMA specifications are used to forecast daily realized volatility of S&P100 index returns:

\[
ARMA(p,q) : \text{Realized}_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \text{Realized}_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t \quad (27)
\]

We use the lag polynomials up to 5 days (or 1 week; \(p = q = 1–5\)) and the results are robust across different lag specifications.\(^{24}\)

---

\(^{24}\)The results from the ARMA\((p,q)\) models are very similar. Following the referee’s recommendation, we report the adjusted-\(R^2\), TIC, HMAE, and HRMSE values only for the ARMA\((1,1)\) model. The results for alternative ARMA models are available upon request.
The adjusted-$R^2$ values and the TIC, HMAE, and HRMSE measures in Table 5 indicate that the in-sample performance of the ARMA model is very similar to that of EVT and VIX, but it performs much better than the GED TS-GARCH model. A notable point in Table 5 is that the 1-day-ahead and especially 20-day-ahead forecasts indicate that the ARMA model cannot outperform the conditional EVT. Although not presented here, the adjusted-$R^2$ values for different ARMA models are in the range of 45% for 1-day-ahead and 24% for 20-day-ahead forecasts, whereas the corresponding values for EVT are about 50% for 1-day-ahead and 37% for 20-day-ahead forecasts. TIC and HMAE measures presented in Table 4 and Panels B–C of Table 5 indicate that there is no clear evidence whether EVT or ARMA model performs better in predicting the 1-day-ahead and 20-day-ahead realized volatility. However, the HRMSE measures for EVT are smaller than those for the ARMA model.

These results suggest that the information content of extreme intraday returns is such that EVT performs well when compared to the forecasts in Eq. (27), which are based on aggregated squared 5-min returns, and thus exploit the information in the entire series of intraday returns.

### 6.2. Unconditional extreme value volatility estimator

The adjusted-$R^2$ values and the TIC, HMAE, and HRMSE measures in Table 5 indicate that the in-sample performance of the ARMA model is very similar to that of EVT and VIX, but it performs much better than the GED TS-GARCH model. A notable point in Table 5 is that the 1-day-ahead and especially 20-day-ahead forecasts indicate that the ARMA model cannot outperform the conditional EVT. Although not presented here, the adjusted-$R^2$ values for different ARMA models are in the range of 45% for 1-day-ahead and 24% for 20-day-ahead forecasts, whereas the corresponding values for EVT are about 50% for 1-day-ahead and 37% for 20-day-ahead forecasts. TIC and HMAE measures presented in Table 4 and Panels B–C of Table 5 indicate that there is no clear evidence whether EVT or ARMA model performs better in predicting the 1-day-ahead and 20-day-ahead realized volatility. However, the HRMSE measures for EVT are smaller than those for the ARMA model.

These results suggest that the information content of extreme intraday returns is such that EVT performs well when compared to the forecasts in Eq. (27), which are based on aggregated squared 5-min returns, and thus exploit the information in the entire series of intraday returns.

### Table 5

Empirical performance of ARMA(1,1) realized volatility model

<table>
<thead>
<tr>
<th>Model</th>
<th>Adjusted-$R^2$</th>
<th>TIC</th>
<th>HMAE</th>
<th>HRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. In-sample performance of ARMA(1,1) model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>47.11%</td>
<td>17.80%</td>
<td>24.44%</td>
<td>35.03%</td>
</tr>
<tr>
<td>Panel B. Out-of-sample performance of ARMA(1,1) model (1-day-ahead forecasting)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>43.93%</td>
<td>18.66%</td>
<td>25.71%</td>
<td>38.60%</td>
</tr>
<tr>
<td>Panel C. Out-of-sample performance of ARMA(1,1) model (20-day-ahead forecasting)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>22.56%</td>
<td>22.48%</td>
<td>30.68%</td>
<td>45.39%</td>
</tr>
</tbody>
</table>

This table shows the in-sample and out-of-sample performance of ARMA(1,1) realized volatility model. Adjusted-$R^2$ values indicate the proportion of the total variation in daily realized volatility that can be explained by ARMA(1,1) model. Theil Inequality Coefficient (TIC) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) measure how far the estimated volatilities are away from the realized volatility.
To clarify this issue, we compare the performance of the unconditional EVT estimator with the conditional EVT in predicting 1-day-ahead realized volatility. We estimate the ‘constant’ location, scale, and shape parameters of the GEV distribution using the past 21 days, 63 days, 126 days, and 252 days of extremes. Then, we use the ‘unconditional’ version of the volatility formula given in Eq. (12b) and compute the unconditional EVT estimator for the next day. For example, we use 21 daily minima from 1/2/1987 to 1/30/1987 and estimate \( m, s, \) and \( x \), and then substitute the estimated scale \( (\sigma) \) and shape \( (\xi) \) parameters into the unconditional volatility formula. Finally, we use this estimator to predict the realized volatility for the next day (2/2/1987). Based on 1-day moving window of model estimation, we obtain a time-series of unconditional EVT estimator. We use the same procedure with an estimation window length of 63 days, 126 days, and 252 days.

As discussed earlier, the relative performance of the conditional EVT estimator is presented based on the minimal returns. To be consistent with the results reported in the paper, we calculate the adjusted-\( R^2 \) values for the unconditional EVT obtained from the minimal returns. The adjusted-\( R^2 \) values are about 28.49%, 35.06%, 21.77%, and 16.22% for 21-day, 63-day, 126-day, and 252-day estimation windows, respectively. They are all smaller than 49.87%, which is the adjusted-\( R^2 \) for the conditional EVT from 1-day-ahead forecast. These results indicate that the conditional EVT performs better than the unconditional EVT in capturing time-series variation in realized volatility.

### 6.3. An alternative assessment of fit

The traditional measure of volatility, based on squared daily returns, has been criticized in the recent literature (see, e.g., Andersen and Bollerslev (1998a) among others) which advocates the use of squared intraday returns as an essentially error-free and model-free measure of volatility. While there certainly are sound reasons to focus on volatility derived from intraday returns, this measure of volatility may have drawbacks also. For example, 5-min returns may be contaminated by microstructure noise. To check whether our results are driven by microstructure noise, we repeated the analysis in Section 5 with 15- and 30-min returns. The results (available upon request) turn out to be very similar to those reported in our tables for 5-min returns.

Also, in certain applications, it may be of interest to forecast volatility based upon squared (or absolute) daily returns, and not integrated volatility. To this end, we also generate a traditional realized volatility measure based on the residuals of the autoregressive of order five process:

\[
R_{t+1} = \alpha_0 + \alpha_1 R_t + \alpha_2 R_{t-1} + \alpha_3 R_{t-2} + \alpha_4 R_{t-3} + \alpha_5 R_{t-4} + \alpha_6 R_{t-5} + \epsilon_{t+1}, \quad (28)
\]
where \( R_{t+1} \) is daily index return, \( E_t(R_{t+1} | \Omega_t) = \alpha_0 + \alpha_1 R_t + \alpha_2 R_{t-1} + \alpha_3 R_{t-2} + \alpha_4 R_{t-3} + \alpha_5 R_{t-4} + \alpha_6 R_{t-5} \) is the daily conditional mean.\(^{26}\) We run an OLS regression to generate the daily realized volatility.

Table 6 presents the in-sample and out-of-sample performance of alternative volatility models in forecasting the traditional measure of realized volatility. The adjusted-\( R^2 \) values indicate the proportion of the total variation in daily realized volatility that can be explained by alternative volatility models. Theil Inequality Coefficient (TIC) always lies between zero and one, where zero indicates a perfect fit. The heteroscedasticity adjusted mean absolute error (HMAE) and root mean squared error (HRMSE) measure how far the estimated volatilities are away from the realized volatility.

\(^{26}\)The optimal lag length in Eq. (28) is determined based on the standard deviation of residuals from AR(p) specifications (where \( p = 1–5 \)) as well as the Akaike Information Criterion (AIC).

\(^{27}\)At an earlier stage of the study, we assess the in-sample and out-of-sample performance of alternative ARMA(p,q) models in forecasting the traditional measure of realized volatility. Since the results from different ARMA specifications turn out to be very similar, we report the adjusted-\( R^2 \), TIC, HMAE, and HRMSE values only for the ARMA(1,1) model.
noisy estimators of day-by-day fluctuations in index returns. Therefore, ARMA(1,1) model performs worse than EVT, VIX, and GED TS-GARCH models in predicting the traditional measure of realized volatility. However, as discussed earlier, when daily realized volatility is measured by high-frequency data, the performance of ARMA(1,1) model is similar to that of EVT and VIX, and it outperforms the GED TS-GARCH model.

6.4. Dependence structure of intraday returns

As pointed out in Section 2, intraday returns are not i.i.d. and, consequently, the limit distribution of extremes need not belong to the domain of attraction of the three standard extreme value distributions (Frechet, Weibull or Gumbel). We verify that our results are not driven by a misspecified conditional extreme value distribution by following Diebold et al. (1998) in conducting the following robustness check: we first fit a conditional mean-volatility model to the raw intraday returns, standardize the data by the estimated conditional mean and volatility, and then repeat our analysis based on these standardized residuals.

We find that an AR(1) GJR-GARCH(1,1) model with GED-distributed errors produces standardized intraday returns that come closest to being i.i.d. (as measured by the correlograms of standardized 5-min returns and squared standardized 5-min returns). We use these standardized 5-min returns to obtain the maxima and the minima for each day, and again estimate the time-varying location, scale and shape parameters of the GEV distribution by maximum likelihood. We then study the relative performance of EVT for 1-day and 20-day ahead forecasts of ‘realized volatility’ (here, the sum of squared standardized 5-min returns). Note that we cannot compare the relative performance of VIX with EVT and GED TS-GARCH models in this context because VIX is the implied volatility of S&P 100 index returns, not the volatility of standardized returns. Since we do not expect VIX to predict the realized volatility of standardized returns, we only compare the in-sample and out-of-sample performance of EVT and GED TS-GARCH models.

Table 7 presents the adjusted $R^2$, TIC, HMAE and HRMSE measures. The results indicate that EVT performs much better than the GED TS-GARCH model in forecasting the volatility of standardized returns.

7. Conclusions

This paper models the conditional distribution of extreme high-frequency index returns, and introduces a conditional EVT based on the GEV distribution. This is also a first attempt towards detecting any time-series variation in extreme value distributions using high-frequency intraday data. The relative performance of the EVT is compared with the discrete-time GARCH and IV models for 1-day and 20-day-ahead forecasts of volatility. Overall, the adjusted $R^2$ measures and the related test statistics imply superior performance of VIX and EVT in capturing time-series
variation in volatility. The forecasting ability of discrete-time GARCH models with GED, Student-\(t\), and normal distributions turns out to be inferior to VIX and EVT.

Although the adjusted-\(R^2\) values provide the direction and magnitude of the relationship between the realized and estimated volatilities, they do not measure how far the volatility forecasts are away from the realized volatility. When the summary statistics are calculated based on the deviation between forecasts and realizations, the relative performance of alternative volatility models changes. The results from the TIC, HMAE, and HRMSE all highlight the superior performance of the EVT against the implied and GARCH volatility models.

Numerous interesting questions for future research remain. Of particular interest to regulators and practitioners, the analysis in this paper has important implications for financial risk management. Previous research has identified the unconditional volatility of extreme returns as an important component of VaR calculations. The conditional extreme value approach advocated here can be readily applied to produce dynamic VaR estimates. In other words, one can propose a conditional extreme value approach to estimating daily VaR based on the time-varying location, scale, and shape parameters of the GEV distribution. This is important, because financial institutions calculate and monitor their VaR on a daily basis, but current extreme-value based methods are unconditional and thus cannot produce VaR estimates that explicitly depend upon current economic conditions.

However, most VaR applications to a portfolio of financial assets require a multivariate version of the conditional approach. In other words, one may have to
distinguish the univariate measures of conditional VaR from the bivariate measures of conditional VaR. Although the existing literature provides a set of bivariate extreme value distributions (e.g., Marshall and Olkin (1983), Tawn (1988), and Joe et al. (1992)), extending our estimation methodology to the conditional bivariate extreme value distributions is beyond the scope of this paper.

Acknowledgements

We thank Carl Chiarella (the editor) and two anonymous referees for their extremely helpful comments and suggestions. We benefited from discussions with Robert Engle, Stephen Figlewski, Roni Michaely, Salih Neftci, Kenneth Singleton, Bhaskaran Swaminathan, and Liuren Wu. The suggestions of Linda Allen, Valentina Corradi, Charlotte Hansen, Lin Peng, and Robert Schwartz have been very valuable. Many thanks go also to seminar participants at the 2005 American Economic Association meetings. Turan Bali gratefully acknowledges the financial support from the PSC-CUNY Research Foundation of the City University of New York.

Appendix A. Moments of the GEV distribution

Let \( X \) be a GEV distributed maxima with cdf \( F_X(x) \), where
\[
F_X(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}.
\]
(A.1)

Consider the transformed variable \( Y \) given by
\[
Y = 1 + \xi \left( \frac{X - \mu}{\sigma} \right),
\]
(A.2)
and compute the cdf of \( Y \), \( F_Y(y) \):
\[
F_Y(y) = \Pr(Y < y)
= \Pr \left( 1 + \xi \left( \frac{X - \mu}{\sigma} \right) < y \right)
= \Pr \left( X < \mu + \frac{\sigma}{\xi} (y - 1) \right),
\]
where in (4) we assume that \( \xi > 0 \). Eq. (1) then implies that
\[
F_Y(y) = \exp \left\{ -y^{-1/\xi} \right\},
\]
(A.5)

Now consider the transformation
\[
Z = Y^{-1/\xi},
\]
(A.6)
we have that
\[ Pr(Z < z) = Pr(Y^{-1/\xi} < z) = Pr(Y > z^{-\xi}) = 1 - Pr(Y < z^{-\xi}) = 1 - e^{-z}. \] (A.7)

So \( Z \) has the exponential distribution with probability density function
\[ f_Z(z) = e^{-z}. \] (A.8)

Then the \( r \)th moment of \( Y \) about zero is given by
\[ E(Y^r) = E(Z^{-r}) = \int_0^\infty z^{-r}e^{-z} \, dz = \Gamma(1 - \xi r). \] (A.9)

It follows from (9) that
\[ E(Y) = \Gamma(1 - \xi), \] (A.10)

and
\[ \text{var}(Y) = E(Y^2) - (E(Y))^2 = \Gamma(1 - 2\xi) - [\Gamma(1 - \xi)]^2. \] (A.11)

From (2) we have that
\[ X = \mu + \frac{\sigma}{\xi}(Y - 1). \] (A.12)

Eqs. (A.10) and (A.12) imply that
\[ E(X) = \mu + \frac{\sigma}{\xi}(E(Y) - 1) = \mu + \frac{\sigma}{\xi}(\Gamma(1 - \xi) - 1), \] (A.13)

and Eqs. (11) and (12) imply that
\[ \text{var}(X) = \frac{\sigma^2}{\xi^2} \text{var}(Y) = \frac{\sigma^2}{\xi^2} \left\{ \Gamma(1 - 2\xi) - [\Gamma(1 - \xi)]^2 \right\}. \] (A.14)


